Handwriting and Pattern Recognition*

MURRAY EDEN†, MEMBER, IRE

Summary—**Handwriting can be characterized as a sequence of basic strokes connected according to rule. Handwriting so generated approximates that of humans very closely. Such a matching process can be used as the fundamental principle in a handwriting recognizer.**

Motivation

THE DESIGNATION, "pattern recognition" has come to be applied to a diversity of problems. The methods employed for these problems frequently do not appear to be related to one another. A good measure of this lack of coherence is, in all probability, attributable to the difference in motivation which characterize problems going by this name. Three quite different motives can be recognized:

1) To construct a procedure to make particular identifications better than a human being, or to make identifications a human cannot make without augmenting his senses by some mechanisms.

2) To make a machine perform certain acts of identification—in general, perceptions which humans can also make -in which the initial structure of the machine and its input is made as simple as possible.

3) To explain some human (or other organismic) act of identification; that is, perception, by means of an appropriate theory or simulation that takes into account not only the act of perception itself, but also what is known about the biological facts of the case.

Developments prompted by the first motive are perhaps the most common. Usually these involve the detection of a signal in the presence of noise. In this case there is little point in asking how humans perform the same task since the goal is to optimize the technological solution rather than to study human behavior.

Such detection techniques are most successful in dealing with signals having a finite set of well defined basis elements. These, perhaps modified by some affine transformations, are to be identified in the presence of noise. The decision procedures in use generally involve the computation of certain statistics of the sample.

The Morse decoder developed by Gold[1] may be considered an example, in this case more likely to be called "pattern recognition" than "signal detection," but embodying the same principles. The basic symbols are surely well defined, variability being introduced in the durations of dots, dashes, and the spaces between symbols. These time variations were studied statistically for a given sender and for a population of senders. For example, Gold obtained the density distribution function for interdot time and interdash time and correlation coefficients relating the several measurable quantities.

One might conjecture that such a device operates a good deal like a human Morse code reader. This is difficult to disprove, but there is evidence that casts doubt on this belief. Thus, suppose the Morse code were being generated by a teletypewriter. Then the machine reader would do as well in identifying English, as French, German, or Hungarian. It is quite doubtful that a human reader who knew only English could do as well with the foreign languages although the letters are the same.[2]

Research influenced by the second motivation has generally concerned models of learning.[3] The handicaps imposed by insisting that the initial organization of the machine be as nearly unstructured as possible and that it "learn" from experience has limited its application to the simplest kinds of "perception." Where such devices have been used successfully on more complicated tasks, for example block letter recognition, it has been necessary to introduce normalization before the learning stage to reduce variations in size and orientation. Thus, the random "learning" network simply compiles statistical data on a small number of nearly fixed patterns.[4]

Thus, a large part of the research effort in pattern recognition has been directed either to pragmatic ends or to general theories of learning or brain organization. While recognition of handwriting surely has practical utility, our principal motivation in studying the problem is third on the list above. A computer was used simply because the computations would have been too long to do by hand. If anything was learned in this experiment, we ourselves and not the machine learned.

So far our principal results have related to the process

* Received by the PGIT, October 9, 1961; revised manuscript received, October 13, 1961. This work was supported in part by the U. S. Army Signal Corps, the Air Force Office of Scientific Research, and the Office of Naval Research; and in part by the M.I.T. Computation Center, the National Science Foundation, and the National Institutes of Health.

† Dept. of Elec. Engrg., Mass. Inst. Tech., Cambridge, Mass.

[1] B. Gold, "Machine recognition of hand-sent Morse code," IRE TRANS. ON INFORMATION THEORY, vol. IT-5, pp. 17–24; March, 1959.

[2] Gold is well aware of the difficulties. He comments, "The conclusion is inescapable, therefore, that for the automatic reception of a language encoded by even a simple process like Morse code, a machine must have some knowledge of the language if it is to approximate the performance of a man."

[3] It is questionable that these pattern recognition devices are relevant in any serious way to human mechanism. The notion that the child's brain begins as a collection of essentially random connections and is organized under the influence of learning, (where are the rules for learning stored? Surely not in random connections.) has as little anatomical justification as has the notion of mind as *tabula rasa* of an earlier psychology.

[4] L. G. Roberts, "Pattern recognition with an adaptive network," 1960 IRE INTERNATIONAL CONVENTION RECORD, pt. 2, pp. 66–70.

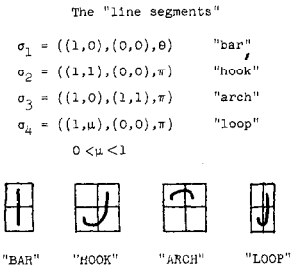

The "line segments"
$\sigma_1 = ((1,0),(0,0),\theta)$ "bar"
$\sigma_2 = ((1,1),(0,0),\pi)$ "hook"
$\sigma_3 = ((1,0),(1,1),\pi)$ "arch"
$\sigma_4 = ((1,\mu),(0,0),\pi)$ "loop"
$0 < \mu < 1$
"BAR" "HOOK" "ARCH" "LOOP"

Fig. 1—A representation of the primitive elements of cursive writing.

of generating handwriting.[5] The justification for investigating the structure of the generative process is treated elsewhere in this issue.[6] Briefly, however, our approach is recognition through a generation process; other approaches to recognition, using analytic techniques, have been described by Frishkopf and Harmon.[7] One aspect of the generative approach; namely, its relation to the definitions of the set of patterns, is worthy of note here.

The customary paradigm for testing if a pattern has been recognized can be described briefly as follows. Suppose the object is to test a program for letter recognition. First, a subject is asked to write a letter. It is then inserted in some appropriate format into the computer. The computer output consists of some symbol standing for one of the finite set of letter equivalence classes. Finally, the subject is asked whether the identification is correct or, more usually, the operator of the experiment determines whether the computer's output agrees with the subject's input. Actually, the experimenter must make two subjective judgments. He must decide what symbol the subject inserted and whether the computer identified it correctly. Now, usually the subject and the experimenter agree, but this need not be so. Even if some convention is adopted—say the initial identification is left to the subject—the set of input patterns can hardly be called well-defined. For instance, a subject might write from right to left and produce "mirror-image" letters.[8] By the convention, these input patterns must be admitted as legitimate candidates for identification even though they have peculiar characteristics introduced by the mirror inversion. We know that there is a quite simple rule transforming these particular patterns into others that are more acceptable.

Identification by a generative procedure leads to a clear definition of the set of permissible patterns. The class of acceptable patterns is simply that set which can be generated by the rules operating on the primitive symbols of the theory.

[5] M. Eden and M. Halle, "The characterization of cursive writing," in "4th London Symposium on Information Theory," C. Cherry, Ed., Butterworth, London, England; 1961.
M. Eden, "On the formalization of handwriting," *Proc. of Symposia in Appl. Math.*, Amer. Math. Society, Providence, R. I., vol. 12, pp. 83–88; 1961.
[6] M. Halle and K. N. Stevens, "Speech recognition: A model and a program for research," this issue, pp. 155–159.
[7] L. S. Frishkopf and L. D. Harmon, "Machine reading of cursive script," in "4th London Symposium on Information Theory," C. Cherry, Ed., Butterworth, London, England, pp. 287–299; 1961.
[8] The writing would appear the same as that reflected in a mirror placed at the left end of a line of conventional writing.

Formalization of Cursive Writing

The first step was to formalize the description of cursive writing. Since this formalization has appeared elsewhere,[5] only a general description will be given. A set of four primitive symbols were defined (Fig. 1), each primitive symbol being a point pair, partially ordered. That is, two points are ordered one above the other or to the right or both, and a sense vector is specified for the tangent to the continuous line to be drawn between the two points.

Two conventions were introduced. These specified that the rotation of each tangent vector from one end of a stroke to the other be 180° and be monotone. That is, the line connecting the two points of a segment must contain no point of inflection. A set of 11 symbols is generated by rotating and reflecting these primitives. Finally a set of 33 strokes is obtained by first allowing each symbol to be translated "up" or "down" (the motivation here is to distinguish, for example, between "*l*" and "*e*"), and second adopting an automatic rule specifying which of the two end points is to be regarded as the first point of the stroke.

With this much apparatus it is possible to define each letter of the language as a unique, finite sequence of strokes, Fig. 2. In this construction, *letters* are regarded as abstract symbols. It was decided to reserve the term *word* for a real object on the paper.

Additional rules for collating the strokes are required. One rule concerns strokes within a single letter, another strokes between letters. Fig. 3 shows these rules applied to the strokes of the word "globe." The rules are obligatory and automatic. Since the letters are already defined by a unique sequence of symbols, the collation rules (which put certain constraints on the way each stroke is placed on the paper relative to its neighbors) may be regarded as redundant. A single ligature rule constrains certain of the geometric properties of the line joining one stroke with the next.

This class of representations for any particular word includes all those regarded as legible copies of this word. However, this formalization does not provide an algorithm for generating a standard (say, Palmer) handwriting and *a fortiori* it provides no explicit procedure, or even suggestions, as to how to generate a match to a particular handwritten representation. In order to approach either of these two results, some functional representation of the line segments corresponding to each stroke and ligature is required.

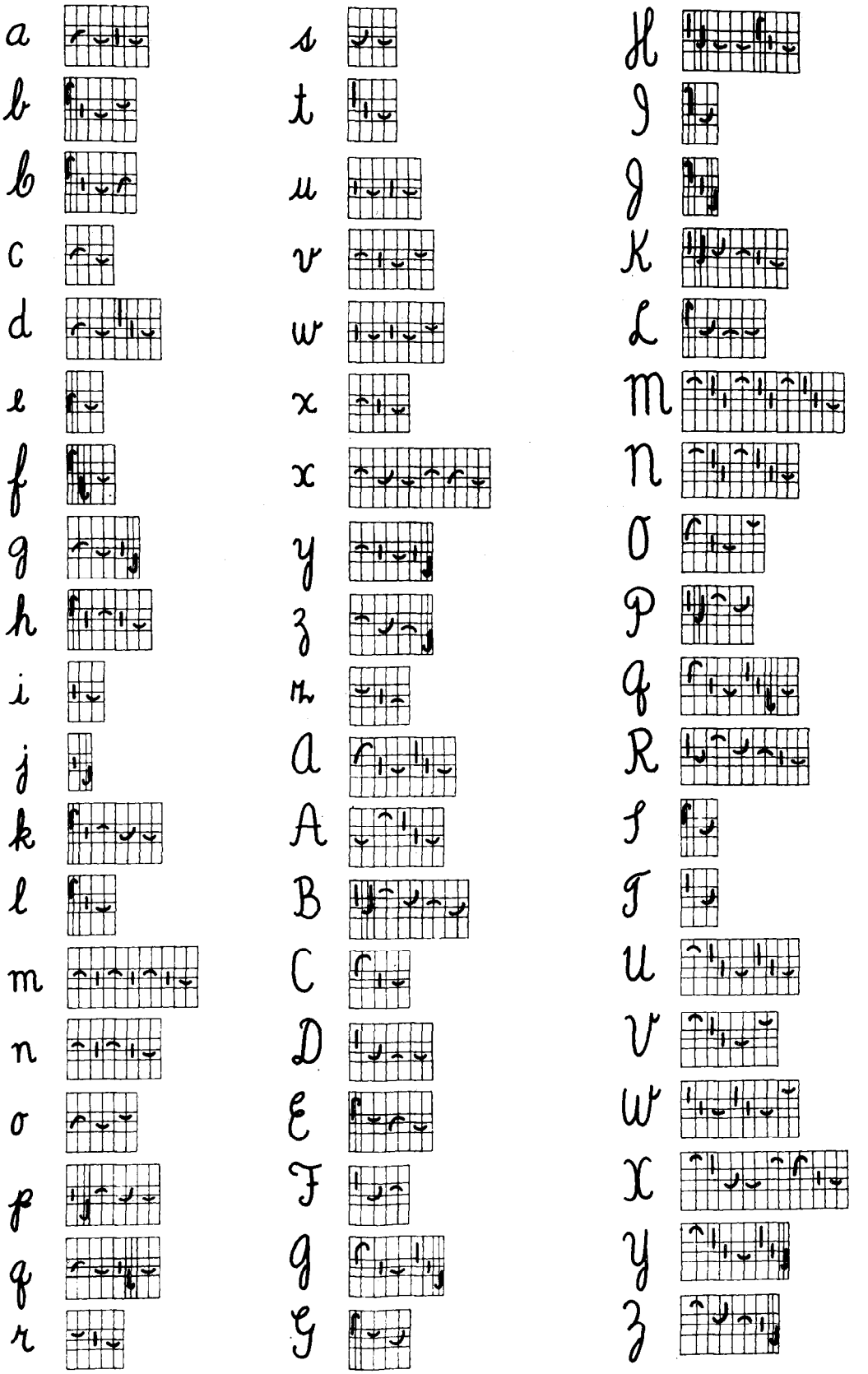


Fig. 2—The sequence alphabet.

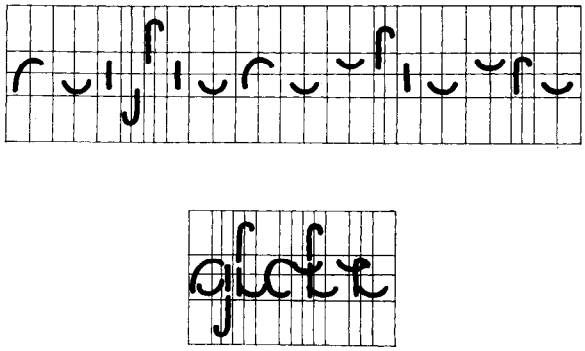


Fig. 3—The stroke sequence of "globe" and a representation of the strokes following the collation rules.

Generation of Cursive Writing

The first attempt to generate handwriting used a random walk in a square plane array; the random walk was constrained by the rules of the theory sketched above. This Markow process can be illustrated by going through the first few steps in generating a replica of the first stroke in the word "globe." It will be noted that the first stroke is a transform of the "hook;" the rules specify that the initial direction of the tangent vector is "up." It follows that the first step of the random walk is to the lattice point directly above the starting point. The rules specify further that the rotation of the vector be counter clockwise. The only admissable next points are then those in the left half-plane. For our purposes it was sufficient to assign zero probability to all points except those marked by an ⊗ in Fig. 4. These points were assigned weights according to an *ad hoc* probability distribution. After the second step was made (Fig. 5), the process could be continued. Now, however, the orientation of the tangent vector being different, the candidates for the next step in the walk are also different. The walk is continued until some point is located directly beneath its predecessor. If this configuration is regarded as defining a stopping rule, then generation of the word is continued by following the rules specified by the next stroke which in this example is a transform of the "arch."

Various probability distributions as well as various stopping rules were used. The writing generated showed considerable variation; its qualitative character is shown in Fig. 6. The words are entirely legible, but they lack the fluidity of a practiced penman. They have been variously described as being childlike, about equivalent to a 8 or 9 year old, or like the writing of a victim of Parkinson's disease. This characteristic is not a consequence so much of the lattice grain or the particular distributions, but rather of the Markow process itself.

There is insufficient coherence or long-range constraints in the line so generated. The shape of the stroke ought to be determined over a longer distance than one lattice interval. This conjecture is consistent with the observation that skilled writers produce about six strokes (and ligatures) per second. This speed is too fast for any detailed control over the pen-point. Presumably the motion of the muscle is essentially ballistic. Once the initial

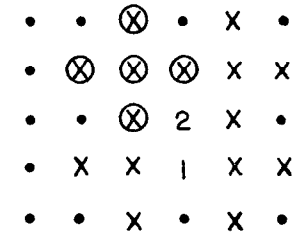


Fig. 4—Admissable lattice points for extension of random walk in the second step.

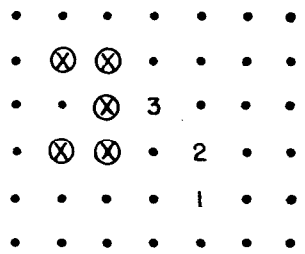


Fig. 5—Admissable lattice points for extension of random walk in the third step.

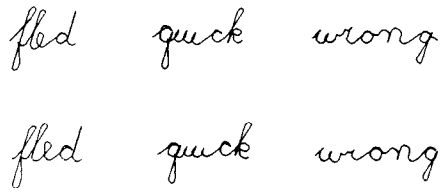


Fig. 6—Words generated by constrained random walk on square lattice.

direction is fixed and forces are applied, the pen-point moves on for a while (say one stroke) before new control signals can come into play.

This conjecture aside, it was felt that information concerning the process might be gained by examining the actual dynamics of writing as a three dimensional function, *time* being the other variable. Accordingly, several records were obtained of y and x displacements as a function of time, Fig. 7. No detailed analysis of these data was undertaken. Qualitatively, it was our impression that the vertical displacement was strongly rhythmic. The horizontal displacement is not so strikingly periodic in its original presentation. However, when the function is corrected for the mean velocity in the x direction, then signs of rhythmicity can be seen in this curve as well. Accordingly, we explored whether the strokes might not be described by segments of periodic functions.

Sinusoidal Description of Practiced Writing

As a first approximation we consider the movement of the pen-point controlled by one pair of muscles governing vertical movements, another pair governing the fluctuating horizontal movements, and a third pair, the steady movement of the pen across the page from left to right. Suppose further that each pair of muscles generates a sinusoid in time. We can regard this formulation as a harmonic oscillator theory for the generation of handwriting.

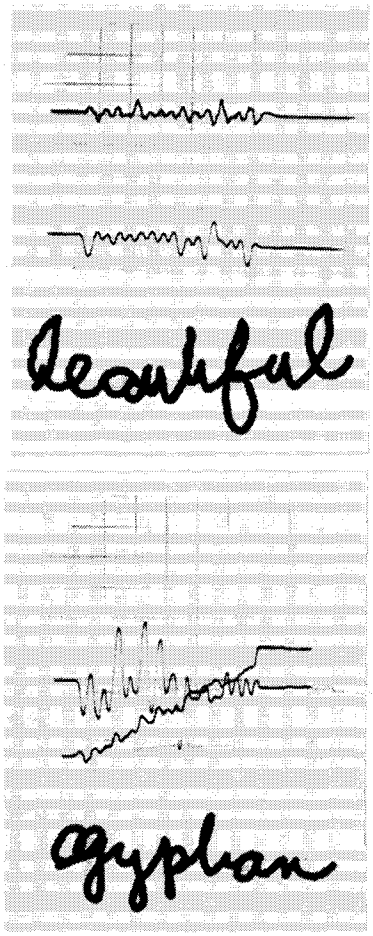


Fig. 7—Oscilloscopic display of handwritten words, heavy trace is x vs t, light trace is y vs t.

The production of the several orthographic exercises in penmanship books can easily be shown to follow from this theory. Thus for example, a series of *l*'s can be produced by appropriate choice of parameters in the equation

$$y = \alpha \sin t$$

$$x = \beta \cos t + vt,$$

where x and y are the spatial co-ordinates, t is time, and α, β, and v are constants. If β is decreased relative to α, a succession of *l*'s (naturally without the crossbar) is obtained. Further, the distance between successive letters is governed by the value of v (in our terminology, the *ramp* parameter). If the sign of β is made negative, then the figure is a succession of m like strokes. If β/α is decreased beyond the critical value, then a figure not found in the standard hand-writing of English is generated. This figure does occasionally appear in rapid writing as the *n*'s begin to be confused with *u*'s. This figure appears also in handwritten Cyrillic as the cursive presentation of λ.

However, the formulation given above does not provide a means for altering the sense of rotation of the tangent vector without introducing discontinuities in the function or its first derivative. This follows from the fact that the two sinusoids have a constant phase relation. If, however, the frequency of one component is made twice that of the other, then the phase relation will change. In particular, if y say, leads by 90° and the frequency of the x component is doubled, then after one-half cycle of the fundamental, y will lag x by 90°. In this way, suitable ligatures containing points of inflection can be generated.[9]

The general form of the stroke equations used in generating handwriting is

$$\dot{y} = \alpha \sin \omega_y t + \nu_y t$$

$$\dot{x} = \beta \cos \omega_x t + \nu_x t \qquad n\pi \leq t \leq (n + m)\pi,$$

where $\dot{x}$ and $\dot{y}$ are the time derivatives of x and y. The physical interpretation of the parameters is fairly straightforward: α is the height of the stroke, β is related to the width of the loop, ν_x is the distance between one stroke and the next. The parameter ν_y is nonzero in a very few letters such as in the majuscule *E*. The equation is written in the derivatives so as to avoid discontinuities in the function. Further, since each stroke is represented by a segment which is an integral multiple of π, the stroke begins and ends parallel to the x axis, and hence there are no discontinuities in the first derivative.

The equation as given above prescribes x and y as orthogonal. Since most writing is slanted, the equation was modified so that instead of plotting x vs y, $y' = y$ was plotted against $x' = x + \sigma y$.

One fortunate (and unexpected) result of this analysis was a further simplification and decrease in the information of an arbitrary sequence. Only the signs of α and β in the initial stroke of the word need be specified, all the other signs being automatically prescribed by the sequence of strokes. Examples of the segments used both for strokes and ligatures are shown in Fig. 8.

Programs for the machine computation of these functions were prepared and various words were "written." The output was displayed as points in a matrix on the face of a cathode ray tube.[10] An example of the output is shown in Fig. 9. It may be considered as an example of a standard handwriting.

Forgery of Cursive Writing

How accurately can the sinusoidal representation replicate actual handwriting? To test this point, words were handwritten on the face of the computer's cathode-ray tube with a "light-pen."[11] This procedure inserted a representation of the word quantized on the 512 × 512 matrix of a 5-inch display. An attempt was then made to fit this input by varying the parameters in the equations, taking each stroke and ligature of the word in the order

[9] It was noted that when practiced, writers were asked to produce a succession of inflection points as in "*n-u-n*" transitions, the speed of writing fell approximately to half. This is reminiscent of the greater difficulty in learning to beat two to the measure in one hand against one to the measure in the other relative to beating both hands in unison.

[10] All the computational experiments were performed on the Dept. of Elec. Engrg. digital computer TX-0.

[11] B. M. Gurley and C. E. Woodward, "Light pen links computer to operator," *Electronics*, vol. 32, pp. 85–88; November 20, 1959.

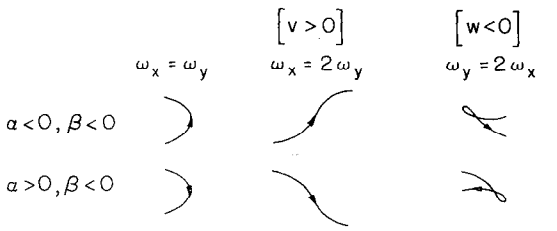


Fig. 8—The function segments defined on the interval $0 \leq th \leq \pi$.

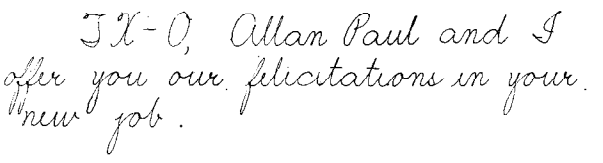
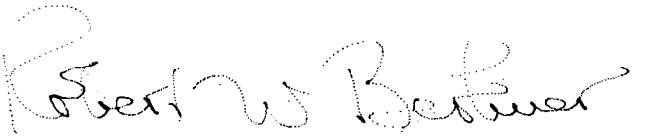


Fig. 9—Display of generated function derived from stroke sequence input and a standard set of parameters.



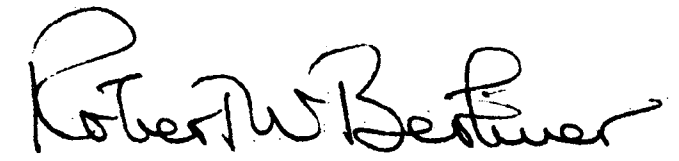


Fig. 10—Example of signature matching: upper signature, machine computed; lower signature, original specimen.

in which it was written. This procedure was used both for signatures, Fig. 10, and for less idiosyncratic writing, Fig. 11. The handwritten sample is below and a machine generated sample above in Fig. 10. In the case of the signatures, matching was approximate since the originals were not written directly on the scope, but were pasted on as a transparency. The handwritten words of Fig. 11 are represented by the sequence of dots that are brighter and more widely spaced, the machine generated words by the lines of denser points.

Judgment of relative goodness of fit was made by eye, and the operator varied the relevant parameters until he was satisfied that he could not easily obtain a better fit. A part, if not all, of the matching procedure can be programmed, but no attempt was made to do so. Examination indicates that qualitatively speaking, the generated signal is very close to the handwritten one.

Although the fit is quite good, it is not necessary that the three dimensional functions themselves agree. However, if the theory presented is to be applied to the physiology of writing then the functions as well as their projections into the plane should match. Fig. 12 illustrates the handwritten and the generated traces for all three projections. The match for $x(t)$ and $y(t)$ is not as close as

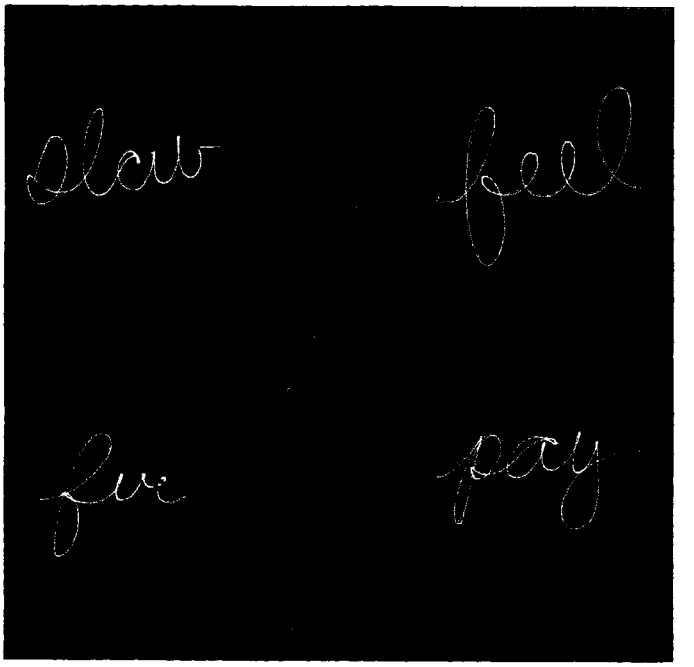
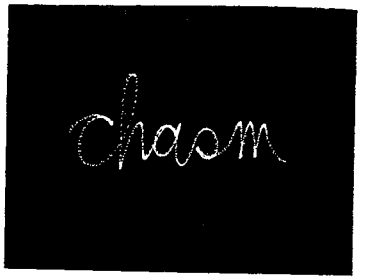
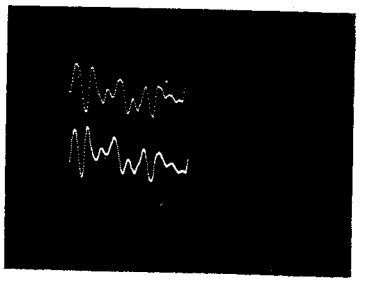


Fig. 11—Four words written by a single subject (widely spaced dots) and the machine computed matches (densely spaced dots).





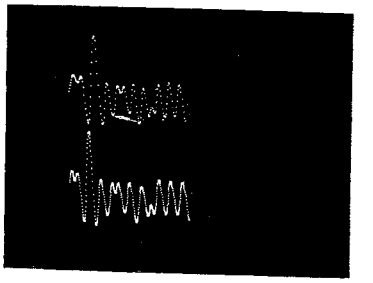


Fig. 12—Top: Display of match of handwritten and generated words. ·····—generated function. · · · · ·—handwriting. Center: Display of $x - \sigma t$ vs t. Upper trace is handwriting; lower trace, generated function. Bottom: Display of y vs t, upper trace is handwriting, lower trace is generated function.

in the x-y plane, but there is still a close resemblance.

Although this is a peripheral issue, the analytical representation of handwritten text provides certain economies in coding the handwriting for transmission. Fixing the grain size of the matrix as described above, five letter words occupied points in a rectangle about 200×500 in extent. Straightforward transmission of the characteristic matrix would require 10^5 bits. On the other hand, specification of the sequence of strokes by their parameters would require about 200 bits, although even this might be decreased by a factor of 4 if an extended text were coded efficiently.

These functions permit investigation of idiosyncrasies in the handwriting of a particular subject. Since each word is defined by the abstract stroke sequence, these idiosyncrasies must depend upon the parameter values of the words written by the subject.

Accordingly, one subject was asked to write a number of different words with the light-pen. Each one was matched and the parametric values recorded. In all, some twenty words were treated. The parameter values obtained may be used as sample populations and their statistical properties studied.

It is very likely that the parametric values of a given stroke, as well as those of adjacent strokes, are not independent. However, as a first approximation, parameter values for each stroke type were plotted as a set of histograms. One of these is shown in Fig. 13. These density distributions were smoothed somewhat and used as weights in selecting the appropriate parameter values in the sequence of strokes used by this subject in writing the word "lay." Fig. 14 illustrates four writings of this word. Three are machine generated versions, the fourth is a handwritten word. It is not immediately obvious which was handwritten and which machine written.[12]

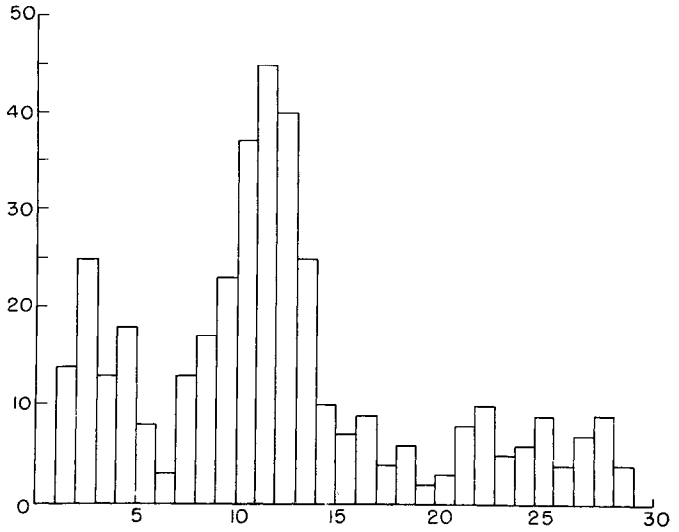


Fig. 13—Histogram of value of the parameter x in the parametric equation for the function segments corresponding to the segments in e, l or u, etc. Units are arbitrary.

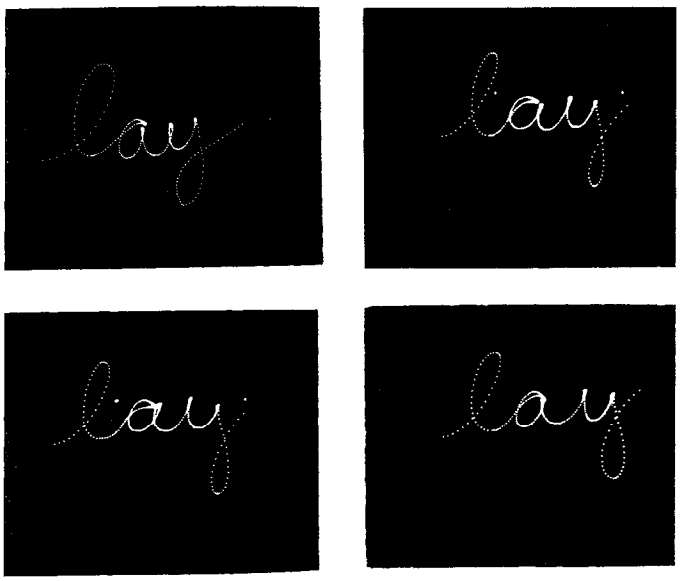


Fig. 14—Three machine generated specimens of the word "lay" and one handwritten word (converted to the same display format).

Handwriting Recognition

The theory of handwriting generation has certain implications for handwriting recognition. The first formalization implies that a word cannot be partitioned into either a unique sequence of strokes or letters. As the handwriting degenerates from textbook script to idiosyncratic form, the number of such ambiguities increases rapidly. If any handwriting reading program is to achieve a score of successes of the same order as that of human readers, some provision must be made for resolving ambiguities by the use of word dictionaries. For sufficiently degenerate handwriting, correct identification is well-nigh hopeless, since there are handwritten texts that can only be read by humans with a knowledge of the syntax of the language and the particular subject matter.

However, the generative theory also suggests that for good handwriting (and to a lesser extent for poor handwriting) the sequence of function segments—as opposed to the sequence of formal strokes—can be determined uniquely. The problem then becomes one of mapping the sequence of function segments into a sequence of strokes under the constraints imposed by the word dictionary. It remains to be seen whether such a program can be successful in practice.

[12] The upper left word in Fig. 14 was handwritten.

Acknowledgment

The work reported here has been conducted in close collaboration with Morris Halle of the Department of Modern Languages, and with Allan Paul and Alan Womack, students in the Department of Electrical Engineering, who have programmed the computations described in the text.